Master's Thesis

Exploring the Relationship between the News and Housing Market Trends: The Sentiment Analysis of Housing Market News in South Korea

Jungseok Seo

Department of Urban and Environmental Engineering
(Urban Infrastructure Engineering)

Graduate School of UNIST

2018

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Exploring the Relationship between the News and Housing Market Trends: The Sentiment Analysis of Housing Market News in South Korea

Jungseok Seo

Department of Urban and Environmental Engineering
(Urban Infrastructure Engineering)

Graduate School of UNIST

Exploring the Relationship between the News and Housing Market Trends: The Sentiment Analysis of Housing Market News in South Korea

A thesis/dissertation
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Jung-Seok Seo

6. 12. 2018

Approved by

Jeongseob Kim

Advisor

Jeongseob Kim

Exploring the Relationship between the News and Housing Market Trends: The Sentiment Analysis of Housing Market News in South Korea

Jungseok Seo

This certifies that the thesis/dissertation of Jungseok Seo is approved.

2018. 6. 12

signature

Advisor: Jeongseob Kim

signature

Gi-Hyoug Cho

signature

Jibum Chung

Abstract

This paper applies text-mining techniques to quantify the tones of news articles. Samples of articles are collected from the most popular website in South Korea, NAVER and the number of articles is 1,299 from Jan 2011 to Mar 2018. Bases on the headlines of articles, a sentiment dictionary for a housing price is constructed by analyzing the relationship of words. Then, it is used to calculate a sentiment index which represents the tone of each news articles. The study shows that the sentiment index has not only a similar pattern with the trends of trading volumes or housing price indices but also some causal relationships with the trends of the housing market. Therefore, the predictive modelings are developed Based on the analysis of time-series. The results show strong explanatory powers for the prices or trading volumes one or two months later. This study is valuable because it doesn't need any costs, but it shows potential for supplementation or substitution of existing consumer sentiment index developed by Korea Research Institute for Human Settlements.

Contents

List of Figures

List of Tables

I. Introduction -- 1
1.1 Background --- 1
1.2 Purpose -- 1

II. Literature Review --- 2
2.1 History of the Predictive Modelling in the Housing Markets --------------------------------- 2
2.2 The Relationship between the Newspapers and Housing Market ----------------------------- 2
2.3 Sentiment Analysis in the News --- 3

III. Research Design --- 4
3.1 Data -- 6
3.2 Sentiment Dictionary -- 10
3.3 Analytical Methods -- 14

IV. Results -- 15
4.1 Descriptive Statistics and Time-Series Patterns of Variables ------------------------------- 15
4.2 The Results of Time-Series Analysis --- 25
4.3 The Results of Predictive Modellings: Autoregressive and Distributed Lags (ARDL) --- 31

V. Conclusions --- 34

References

List of Figures

Figure 1. Research Procedure -- 4
Figure 2. Structure of Web Page -- 7
Figure 3. The Example of News Articles and Source Codes in NAVER ---------------------------------- 7
Figure 4. Word Cloud of Nouns in News Headlines -- 7
Figure 5. Word Clouds of Nouns in News Headlines: Left - only positive news and Right - only negative news -- 8
Figure 6. Words Relationships in the Headlines (only positive news) ------------------------------------ 11
Figure 7. Words Relationships in the Headlines (only negative news) ----------------------------------- 11
Figure 8. The Example of Sentiment Analysis --- 14
Figure 9. Trends of Housing Volume of Transaction and Sentiment Index, 2011.7 - 2018.2 ----------- 19
Figure 10. Trends of Unsold Apartments and Sentiment Index, 2011.7 - 2018.2 ------------------------ 19
Figure 11. Trends of Begin Construction of Housings and Sentiment Index, 2011.7 - 2018.3 --------- 19
Figure 12. Trends of Apartment Real Sale Price Index and Sentiment Index, 2011.7 - 2018.2 -------- 20
Figure 13. Trends of Apartment Sale Price Index and Sentiment Index, 2011.7 - 2018.3 -------------- 20
Figure 14. Trends of Apartment Chonsei Price Index and Sentiment Index, 2011.7 - 2018.3 ---------- 20
Figure 15. Trends of Consumer Sentiment Index of Housing Market and Sentiment Index, 2011.7 - 2018.3 -- 21
Figure 16. Trends of Consumer Sentiment Index of Housing Sale Market and Sentiment Index, 2011.7 - 2018.3 -- 21
Figure 17. Trends of Consumer Sentiment Index of Housing Chonsei Market and Sentiment Index, 2011.7 - 2018.3 -- 21
Figure 18. Trends of Housing Volume Transaction (1st Difference) and Sentiment Index, 2011.7 - 2018.2 -- 22
Figure 19. Trends of Unsold Apartment (1st Difference) and Sentiment Index, 2011.7 - 2018.2 ------ 22
Figure 20. Trends of Apartment Real Sale Price Index (1st Difference) and Sentiment Index, 2011.7 - 2018.2 -- 23
Figure 21. Trends of Apartment Sale Price Index (1st Difference) and Sentiment Index, 2011.7 - 2018.3 --- 23
Figure 22. Trends of Apartment Chonsei Price Index (1st Difference) and Sentiment Index, 2011.7 - 2018.3 -- 23
Figure 23. Trends of Consumer Sentiment Index of Housing Market (1st Difference) and Sentiment Index, 2011.7 - 2018.3 --- 24
Figure 24. Trends of Consumer Sentiment Index of Housing Sale Market (1st Difference) and Sentiment Index, 2011.7 - 2018.3 --- 24

Figure 25. Trends of Consumer Sentiment Index of Housing Chonsei Market (1st Difference) and Sentiment Index, 2011.7 - 2018.3 -- 24
Figure 26. The Cross-Correlation between of Housing Volume of Transaction and Sentiment Index: Left - HVT and Right - D_HVT -- 26
Figure 27. The Cross-Correlation between of the Number of Unsold Housings and Sentiment Index: Left - NUT and Right - D_HVT -- 26
Figure 28. The Cross-Correlation between of Begin Construction of Housings and Sentiment Index: Left - BCH and Right - D_BCH -- 27
Figure 29. The Cross-Correlation between of Apartment Real Sale Price Index and Sentiment Index: Left - ARSPI and Right - D_ARSPI -- 27
Figure 30. The Cross-Correlation between of Apartment Sale Price Index and Sentiment Index: Left - ASPI and Right - D_ASPI -- 28
Figure 31. The Cross-Correlation between of Apartment Chonsei Index and Sentiment Index: Left - ACPI and Right - D_ACPI -- 28
Figure 32. The Cross-Correlation between of Consumer Sentiment Index of Housing Market and Sentiment Index: Left - CSIHM and Right - D_CSIHM -- 29
Figure 33. The Cross-Correlation between of Consumer Sentiment Index of Housing Sale Market and Sentiment Index: Left - CSIHSM and Right - D_CSIHSM -- 29
Figure 34. The Cross-Correlation between of Consumer Sentiment Index of Housing Chonsei Market and Sentiment Index: Left - CSIHCM and Right - D_CSIHCM -- 29

List of Tables

Table 1. Data Description -- 5

Table 2. The lists of positive sentiment dictionary --- 12

Table 3. The lists of Negative sentiment dictionary -- 13

Table 4. Summary statistics -- 18

Table 5. The results of autocorrelations and partial autocorrelations ------------------------------------ 25

Table 6. The results of granger causality analysis --- 30

Table 7. The results of predictive models --- 32

I. Introduction

1.1 Background

2008 Global Financial Crisis gives a lesson that a high return always involves high risk and a wrong prediction in macroeconomics causes disastrous results in various areas including the housing market. In South Korea, the government has intervened in the housing market when it is overheated or depressed because a housing is the most important asset and it can affect social welfares (Nassirtoussi et al., 2013). Although the strong regulations had short-term effects, the housing policies have been criticized by belated actions or it often failed to achieve the main goals of the policies. For instance, the governments announced several regulations to control the housing prices of speculation-ridden areas in 2017, but the prices in many areas, such as Gangnam, are still increasing. As a result, the government acknowledged the failure of policies.

One of the methods to make a better housing policy is to develop a predictive modeling for trends of the housing. By predicting future trends, it is possible to make the better housing markets. To do so, it is necessary to examine the perceptions of potential consumers and experts in the markets. So, Korea Research Institute for Human Settlements conduct sample surveys every month and calculate the consumer sentiment indices since Jan 2011. However, it needs large samples and costs to improve the performance of the index.

Recently, as the techniques of Big Data is being developed, it gives some opportunity to quantify psychological factors of consumers from enormous online text data by using the text-mining. And it is used for the market prediction (Nassirtoussi et al., 2013; Schumaker et al., 2012; Yu et al., 2013). However, there are few studies in South Korea because it is more difficult to apply text-mining of Korean than English.

1.2 Purpose

This study analyses unstructured text-based data which is the online news articles through the technique of text-mining. The main goal is to develop a sentiment index which quantifies the tones of newspapers regarding housing prices. To do so, a topic-specific sentiment dictionary is necessary to improve the accuracy. This index is made by sentiment analysis and it is used to develop predictive modelings for the housing markets. In this process, it gives insight into the role of newspapers in the housing market. Hence, the research hypotheses areas in the following:

(1) The newspapers of housing prices have an influence on a decision-making process of potential consumers in the housing market
(2) The tones of news articles about housing prices have correlation or causal relationships with trends of the housing market
(3) The Sentiment Index, which represents the tones, can predict the trends of the housing market

Through the predictive modelings, it is expected to contribute to not only prospect the trends of housing markets but also establish a housing policy. At least, this study shows that there are some relationships between the tone of newspapers and the trends of housing market

II. Literature Review

2.1 Predictive Modelling in the Housing Market

According to Efficient Market Hypothesis, any prediction is impossible in the efficient Market (Fama, 1965). However, Fama (1970) insisted that information asymmetry can decrease the efficiency of the market and it becomes possible to develop a predictive modeling as the market becomes inefficient. The housing market is well known for the existence of information asymmetries between sellers and buyers (Nassirtoussi et al, Aghabozorgi, Wah and Ngo, 2014). Even sellers have usually enough information about only their neighborhoods. Hence, the housing market is considered as the inefficient market and it has high potential to be predicted by quantitative approaches (Nassirtoussi et al., 2014).

Adams (1964) showed that psychological attitudes have the high explanation of housing consumption. He insisted that Consumer's sentiments be one of most significant factors of the predictive model in the housing market.

2.2 The relationship between the newspapers and housing market

Journalists have voices to make the readers support specific positions (Kam and Song, 2012; Munro, 2018). Although expressions may be not explicit, newspapers can have hegemonic power to restrain ability to think by oneself and have biased ideology supported by elite groups (Munro, 2018). For example, Fairclough (2000) showed that passive voice, about globalization in policy documents, lets people think globalization to be taken for granted. Hence, it is essential to examine the detailed structure and language of the text, which contains a distinct writer's intentions and can influence the psychological factors of the public (Munro, 2018).

The housing market is also closely interacted with newspapers. Newspapers deliver not only the facts such as the trends of the housing market but also they can affect on the market by influencing people's behaviors (Robertson, Geva, & Wolff, 2006; Wisniewski & Lambe, 2013). In addition, perceptions of same information differ depending on past experiences and present conditions (Friesen and Weller, 2006)

Recently, online newspapers are much more common, and it has become easier to collect online articles of a certain topic with a specific keyword (Kam and Song, 2012). In addition, as text-mining is being developed, it becomes possible to analyze the text-based big data of news articles (Kam and Song, 2012). Compared to other online sources, the text-mining of news articles is easier because news articles have less grammatical errors (Nassirtoussi et al., 2014).

2.3 Sentiment Analysis in the News

Sentiment analysis is a technique of text-mining that it automatically categorizes positive or negative sentiments based on the text (Balahur and Steinberger, 2009; Lee, Cui, and Kim, 2016). In other words, overall sentiment of a new article is determined by words in the article and a sentiment dictionary which predefines polarity of words (Kim and Kim, 2014).

One of the initial projects of sentiment analysis in news is Europe Media Monitor (EMM) developed by The European Commission's Joint Research Centre (Balahur and Steinberger, 2009). EMM has daily news gathering engine that collects approximately 300,000 news articles per day in about 70 languages. By using EMM data, it is attempted to evaluate the public sentiment about a certain topic such as housing policy. The basic idea is (1) to classify news articles with the tone of a certain issue and (2) to find time-series trends on the same topic (Balahur and Steinberger, 2009).

When the sentiment analysis is conducted, there is often difficulty to decide whether a text is positive, negative or neutral (Balahur and Steinberger, 2009). There is one example of Korean news headline: "소득 대비 집값 하락세…강남 '불패'는 여전". While "하락세" indicates a negative sentiment of housing prices, "불패" may give a positive feeling of the sentence. Naturally, language is differently interpreted depending on various contexts (Balahur and Steinberger, 2009; Lee, Cui, and Kim, 2016). Moreover, there are many homonyms. In this regard, it is necessary to develop the subject-specific sentiment dictionary to mitigate the conflicting interpretation and improve the accuracy of the analysis (Lee, Cui, and Kim, 2016).

Especially, text mining for Korean is relatively less developed than for English due to the nature of Korean language. In addition, there are few studies related to the housing market

III. Research Design

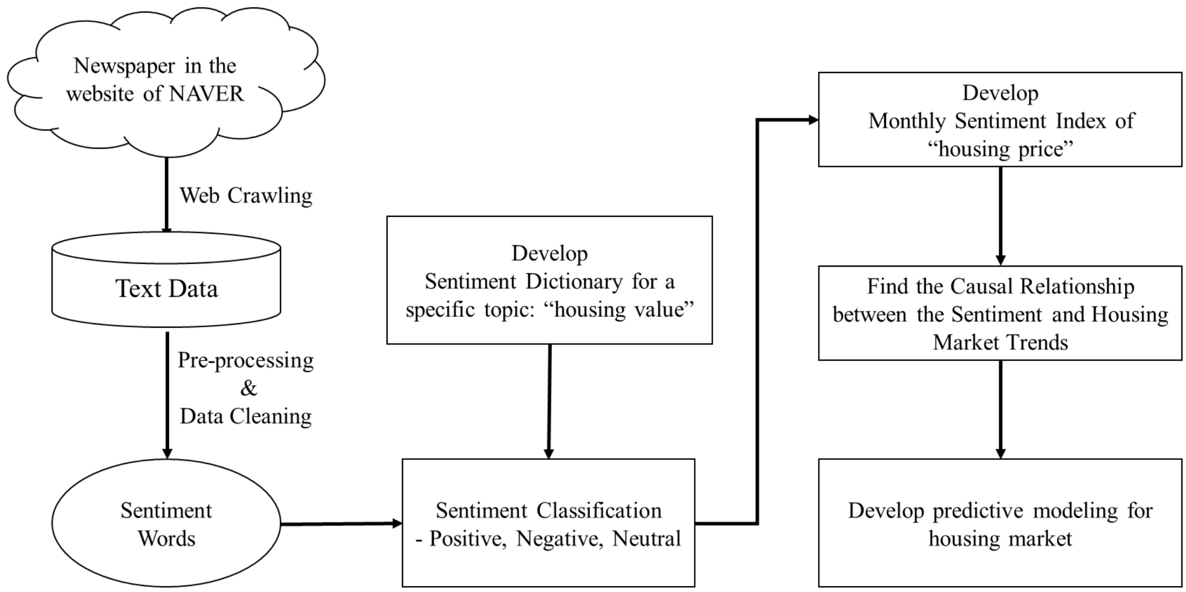

Figure 1. Research Procedure

A research procedure is described briefly in Figure 1. Firstly, samples of newspaper coverage related to the housing prices are collected from the most popular websites in South Korea[1]. This study considers only five major newspaper companies[2], assuming that these can be representative of all newspapers in South Korea. Then, text-mining techniques, such as web crawling and sentiment analysis, are used to develop monthly sentiment index of the housing prices by using news articles. Then, time-series analysis is conducted to find causal relationships between the developed index and housing market trends. Finally, this study tries to establish predictive models for the housing market.

The detail research design is organized as follow: Section 3.1 describes the data. Section 3.2 explains how to develop sentiment dictionary. Lastly, Section 3.3 describes the analytical methods.

[1] NAVER provides news articles from more than 400 media sources

[2] Chosun Ilbo, JoongAng Ilbo, Dong-A Ilbo, Kyunghyang Shinmun, The Hankyoreh

Table 1. Data Description

Variables Name	Description	Source
Sentiment Index (SI)	· The monthly variables which indicate the tones of headlines in newspapers · The ranges of values are from -1 to +1: higher values mean more positive tones regarding housing prices	-
Housing Volume of Transaction (HVT)	· The monthly variables of apartment volume of transaction in Seoul · D_HVT is 1st differential values of HVT	Korea Appraisal Board
The Number of Unsold Apartments (NUA)	· The monthly variables of the number of unsold apartments in Seoul · D_NUA is 1st differential values of NUA	Korea Appraisal Board
Begin Construction of Housing (BCH)	· The monthly variables of the number of begin construction of housings in Seoul · D_ BCH is 1st differential values of BCH	Ministry of Land, Infrastructure and Transport
Apartment Real Sale Price Index (ARSPI)	· The monthly variables calculated by samples of the real sale price of apartments in Seoul · D_ ARSPI is 1st differential values of ARSPI	Korea Appraisal Board
Apartment Sale Price Index (ASPI)	· The monthly variables developed by sample surveys of apartment sale values in Seoul · D_ ASPI is 1st differential values of ASPI	Korea Appraisal Board
Apartment Chonsei Price Index (ACPI)	· The monthly variables developed by sample surveys of apartment chonsei values in Seoul · D_ ASPI is 1st differential values of ASPI	Korea Appraisal Board
Consumer Sentiment Index of Housing Market (CSIHM)	· The monthly sentiment variables developed by sample surveys of four trends: (1) selling and buying; (2) rent and lease; (3) housing sale prices and (4) transaction of housing in Seoul · The ranges of values are from 0 to 200: '> 100' means the positive prospect of the housing market · D_ CSIHM is 1st differential values of CSIHM	Korea Research Institute for Human Settlements
Consumer Sentiment Index of Housing Sale Market (CSIHSM)	· The monthly variables developed by sample surveys of housing sale prices in Seoul · D_ CSIHSM is 1st differential values of CSIHSM	Korea Research Institute for Human Settlements
Consumer Sentiment Index of Housing Chonsei Market (CSIHCM)	· The monthly variables developed by sample surveys of housing Chonsei prices in Seoul · D_ CSIHCM is 1st differential values of CSIHCM	Korea Research Institute for Human Settlements

- 1^{st} differential values indicate the rate of change of original variables

3.1 Data

News Articles

This study analyzes news articles provided on NAVER websites. NAVER builds a database of news from almost all the press and serves news on their web pages. The web pages consist of various components (Figure 2) and text data can be extracted from HyperText Markup Language (HTML) through web scraping. It is the automatic technique to gather and copy specific information from the World Wide Web (WWW).

Web scraping is required to understand a source code of web pages (Figure 3) which indicates where specific text data is located. This study utilizes the free software (called R) for web scraping. R provides various statistical and graphical techniques. Moreover, R can be extended via packages provided by the Comprehensive R Archive Network (CRAN) family of Internet sites. Especially, "httr" and "rvest" packages are used to realize web scraping.

Specifically, news articles of the housing prices are collected from 2011.1 to 2018.3. The major five companies, which have the largest portion of subscribers in South Korea, are included in the study. To identify whether news are related to the topic, three keywords must be included in headlines: "아파트 + 값"; "아파트 + 가격" and "집 + 값".

As a result, 1,299 articles are collected. Each newspaper company wrote averagely 4.4 articles about the housing prices every month. 724 papers were written positively while 462 ones were written negatively. 113 articles had neutral positions on the housing prices. Interestingly, it is realized that the most news articles mentioned Seoul housing markets (Figure 4). It may be because both readers are usually more interested in Seoul than others and Seoul housing market is the most important in the housing market in South Korea. Therefore, this study tries to examine the relationship between news articles and the housing market in only Seoul.

In addition, articles mentioned "강남" and "수도권" and the contents are usually about the trends of housing prices with redevelopment and housing policies (Figure 4 & 5). When word clouds of nouns are conducted in the headlines of article separately, the result in the negative news regarding housing prices shows that the frequency of positive words such as "상승세" and "상승폭" is significantly high. It means that a single word may be the biased indicator of whether an entire sentence has the positive tone or negative tone. In other words, the combinations of words should be considered to determine the sentiment of texts.

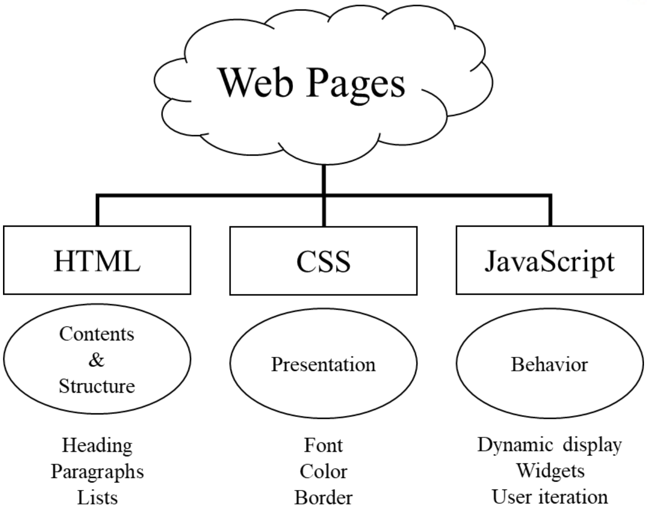

Figure 2. Structure of Web Page

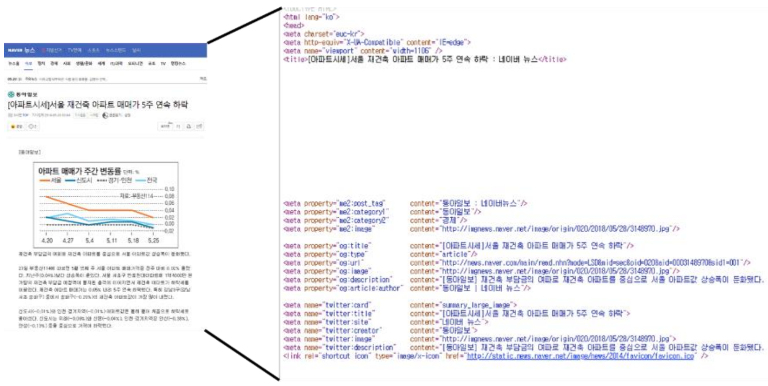

Figure 3. The Example of News Articles and Source Codes in NAVER

Figure 4. Word Cloud of Nouns in News Headlines

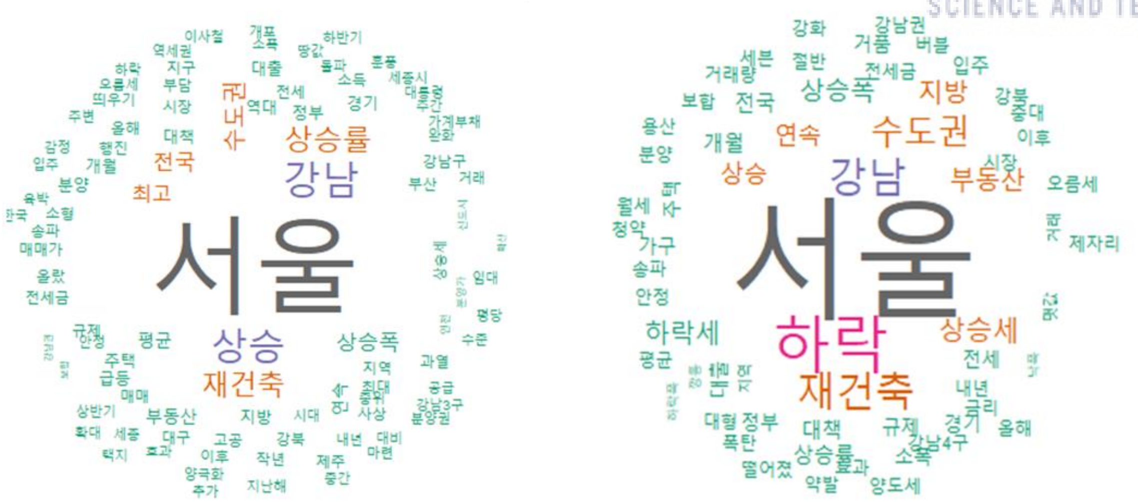

Figure 5. Word Clouds of Nouns in News Headlines: Left - only positive news and Right – only negative news

Variables in the housing market

To examine the effects of tones of newspapers on the housing markets, this study collects nine monthly time-series variables in Seoul housing market (Table 1). The Sources of data are from Korea Appraisal Board, Ministry of Land, Infrastructure and Transport, and Korea Research Institute for Human Settlements. The temporal boundary is from July in 2011 to March in 2018 because consumer sentiment indices have been developed since July in 2011.

Firstly, "Housing Volume of Transaction (HVT)" indicates the number of transactions of apartments in Seoul. Newspapers are the important sources which affect a decision-making process of consumers in the housing markets. If a price of a product increases, the demand of product will decrease according to the law of supply and demand. However, if a price of the product is expected to increase, demand can also increase because people are likely to gain future profit (Engelhardt, 2003). Especially, housings are considered as one of the biggest assets in South Korea. So, people want to buy a new housing of which value will increase in the future (Kim and Yu, 2013; Carroll et al., 1994). Therefore, it is assumed that positive tones about the housing prices in the news increase the volume of transactions.

While "HVT" reflects the demand of the housing market, "The Number of Unsold Apartments (NUA)" and "Begin Construction of Housing (BCH)" are related to supply of the housing market. If housings are oversupplied in the market, it may increase a possibility of unsold apartments (Seo, Lee and Jung, 2010). Moreover, a negative prospect regarding housing prices can cause the increase in the number of unsold apartments. Therefore, it is assumed that negative tones about the housing prices in the news increase the number of unsold apartments. In the case of BCH, positive tones may be correlated positively because a construction company, the actor of construction, pursue to future profit and higher prices of housings indicates more profits by selling constructed housings. Hence, it is assumed that positive tones about the housing prices in the news increase the number of begin construction of housings

The next things are variables of the housing prices. Index of the housing prices should affect the tones about the housing prices in the news because the news articles deliver the fact of trends of housing markets. However, the tones also can influence the index by affecting the decision-making process of consumers. People who expect the upward trend of the housing market are more likely to buy a new housing. So, it is assumed that positive tones about the housing prices in the news increase the index of housing prices.

In this study, three indices are used to represent the housing prices in Seoul. The first variable is "Apartment Real Sale Price Index (ARSPI)". It is calculated by the repeated sale price model. This model includes real trading samples which have more than 2 times transactions in a certain period. By doing so, the index can show the rates of change in the housing prices over times. The second index is "Apartment Sale Price Index (ASPI)". It is developed for showing the trend of the housing market and providing the reference when a housing policy is established. A method and sample of calculation is distinct from ARSPI. It collects samples which can represent a population of housings markets and the index is calculated by the method of Jevons index. The third variable is "Apartment Chonsei Price Index (ACPI)". It is calculated by the same method of ASPI, but it considers only samples of chonsei[3].

Lastly, the consumer sentiment indices, which are developed by Korea Research Institute for Human Settlements (KRIHS), are included in the analysis. It is to identify the change of perception and behavior of consumers. The indices are estimated by sample surveys with 29 questions about the trends of the housing market. It reflects the opinions of both licensed real estate agents and potential consumers. KRIHS provide three types of consumer sentiment indices: (1) Consumer Sentiment Index of Housing Market (CSIHM); (2) Consumer Sentiment Index of Housing Sale Market (CSIHSM) and (3) Consumer Sentiment Index of Housing Chonsei Market (CSIHCM). The ranges of values of each index are 0 to 200. If the value is higher than 100, the perception of the consumer is positive in the housing market.

One of the main purposes of this study is to develop a new sentiment index based on the tones of news articles. To evaluate the performance of the newly developed index, we can compare the existing indices developed by KRIHS. So, it is assumed that both new and existing indices are correlated.

---

[3] Chonsei is the unique methods of trading housings in South Korea. It is the kind of rent and a renter pays rent at once for 2 years. Then, the renter gets back rent when the contract is expired.

3.2 Sentiment Dictionary

A sentiment dictionary is used to determine whether the tones of text such as news articles are positive or negative regarding a certain topic. It is actively developed based reviews of movies because each review has a score and it is possible to evaluate the accuracy of the sentiment dictionary with scores (Lee, Cui and Kim, 2016). However, existing sentiment dictionary has poor performance for other topics (Lee et al., 2016). Therefore, it is necessary to make a topic-specific sentiment dictionary for improving the efficiency.

In this study, headlines of news articles regarding the housing prices are used to develop the sentiment dictionary for the housing market. The headline is the most powerful element of delivering the writer's intention to readers (Munro, 2018). Not only does it include the most important contents, but also it is easy to read and written tastefully (Richardson, 2007). So, analysis of headline is simpler than analyzing entire content, but it is practical to identify the ideological positions (Munro, 2018).

To build the dictionary, this study conducts the analysis of words relationships which shows frequencies of appearance of words simultaneously (Figure 6 & Figure 7). Figure 6 is the results of only positive news of the housing prices while Figure 7 is the results of only negative ones. In the Figures, green lines connect the words which were written in the title of the same news article. The thicker lines are, the stronger words have relationships. Some words such as "껑충" and "거품" don't have any connection. It means that those independent words can be used to determine whether the tone of articles is positive or negatives. On the other hands, "상승" cannot be used solely to decide the tone because it is connected with various words and it can be both negative and positive tone depending on the combinations of words.

By examining the relationships of words, the candidates are selected and tested whether they have consistent results to filter the tones. The candidates can be both single words and combinations of words and if it is checked that they can be representative of positive or negative tones, it is included in the sentiment dictionary. However, there are some articles which don't contain those candidates. In the cases, the representative combinations of words are chosen from each article manually.

As a result, Table 2 shows that the sentiment dictionary which indicates the positive tones of headlines and table 3 shows that the dictionary of the negative ones. There are 91 patterns of positive tones and 90 patterns of negative tones.

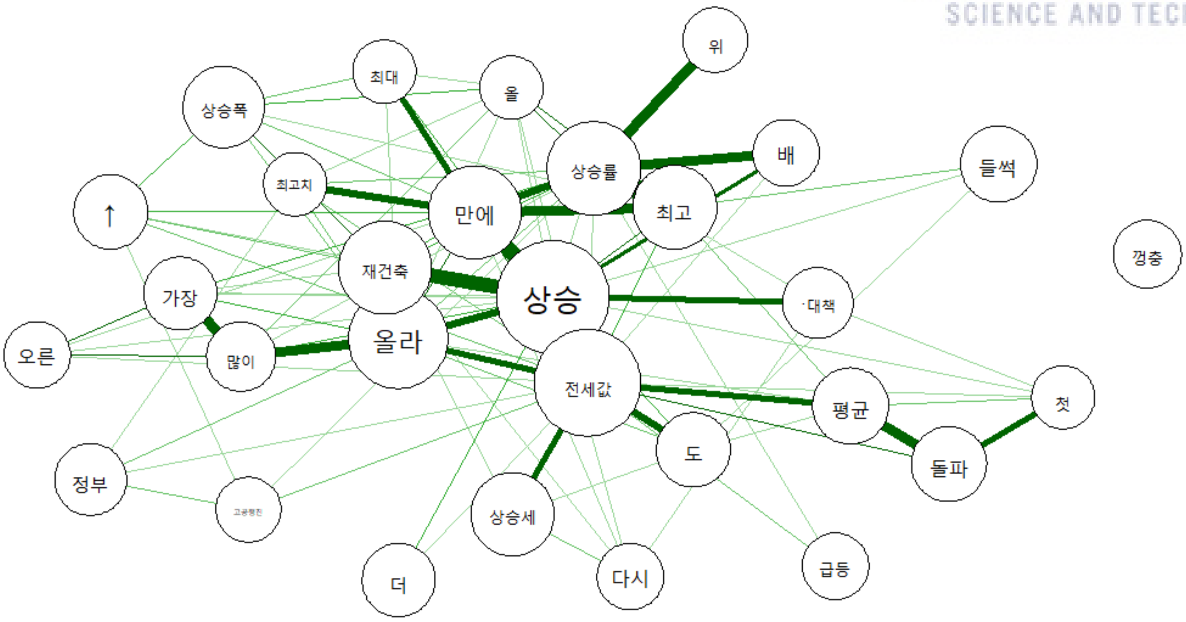

Figure 6. Words Relationships in the Headlines (only positive news)

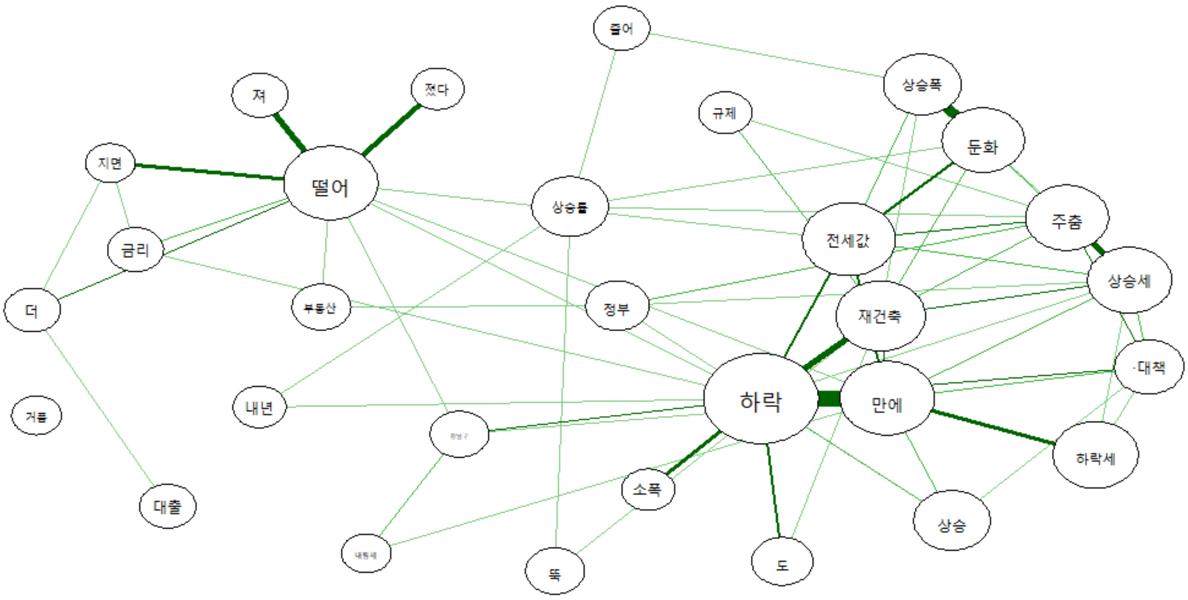

Figure 7. Words Relationships in the Headlines (only negative news)

Table 2. The lists of positive sentiment dictionary

Positive Sentiment Dictionary			
↑	뛰었	수혜	최고치
[^가구] 육박	뛴	숨통 트이나	최대폭 상승
가치 상승	뜀박질	시대	큰.*상승
가팔라	뜨거	쑥쑥	투기 재연
강세	띄우기	앞질러	튀네
갬	만에 상승	약발 끝났나	펄펄
거래.*활발	만에 오름	약발은	풍부한
거품 아니다	많이 올	에도 상승	하락세.*멈
高	무섭게 상승	역대.*최고	하락하던
고개드는	바닥쳤나	역세권 불패	호재
고공행진	반등	연속 상승	활기
과열.*해제	부글부글	연속 오름	활황기
급등	부터 상승	오르나	후끈
급증	불붙은	오를 것	훈풍
껑충	불확실성 해소	오를것	
꿈틀	사상.*최고	오름.*커	
넘어	상승 기대감에	오름세 전환	
높은 주거비	상승 기대에	온기	
누적 상승률	상승 이어져	올라	
다시 상승세	상승.*진입	올랐다	
돌파	상승.*최대	육박	The number of patterns: 91
동반 상승	상승.*커	잡으려	
들썩	상승.*큰폭	재건축	
떨어지던	상승.*확산	주째 상승	
뛰네	상승세 번	지금이 바닥	
뛰는	상승폭 더	천장 뚫	
뛰어	상승폭.*확대	최고.*상승	

[^word]: The sentence starts from the word
A.*B: any words between A and B

Table 3. The lists of Negative sentiment dictionary

Negative Sentiment Dictionary			
↓	다시 하락세	안정세	추락
8 2.*대책	대책 임박	약발 계속	축소
감감	대책 주시	약세	침체
강력 규제	대출 규제	엇갈려	폭락
거래 한산	둔화	움찔	풍선효과.*차단
거래량.*감소	떨군	위기	하락
거품	떨어	위축	하우스푸어
결국.*내리막	뚝뚝	잠잠	한숨
경고	마지노선	잡기	한파
곤두박질	머뭇	잡는다	
과열	멈춰	전세난	
그쳐	멈췄	절반	
그칠	모니터링	정부규제 칼	
금리.*인상	무너	제자리	
급감	미분양 늘	좁	
급락	바닥	주춤	
깎	반토막	줄어	
깡통	변수	줄었	
깨	부정사전	집값.*빠	The number of patterns: 90
꺾	붕괴	짙은	
꽁꽁	빙하기	쪼그라든	
낙폭	상승 멈	찔끔	
내려	상승.*정체	찬바람	
내렸	상승.*제한	천장.*쳤나	
내림	속탄다	최저	
눈치 보기	숨고르기	추가.*규제	
눈치보기	악몽	추가.*대책	

[^word]: The sentence starts from the word
A.*B: any words between A and B

3.3 Analytical methods

Sentiment Index

Sentiment Index (SI) quantifies the tones of headlines in news articles. The equation (1) is used to calculate SI:

$$\text{Sentiment Index} = \frac{\text{The number of positive patterns} - \text{The number of negative patterns}}{\text{The number of morphemes}} \quad (1)$$

To determine either positive or negative patterns in the texts, the developed sentiment dictionary is utilized. This method enables to construct continuous variables of the sentiment. So, it is possible to distinguish more positive (negative) texts among all positive (negative) texts. In addition, this index expresses the trade-offs when a text has both positive and negative patterns.

The example of how to quantify the sentiment of a text shown in figure 8. Firstly, stop words are removed from an original text of headline because stop words are not meaningful. Then, morphological analysis is conducted to distinguish each independent word. Then, the number of positive and negative patterns are calculated based on the sentiment dictionary. In this process, the sentiment index of (“[매매시황] 서울 아파트값 8주째↓… 짙은 관망세”) is -0.333.

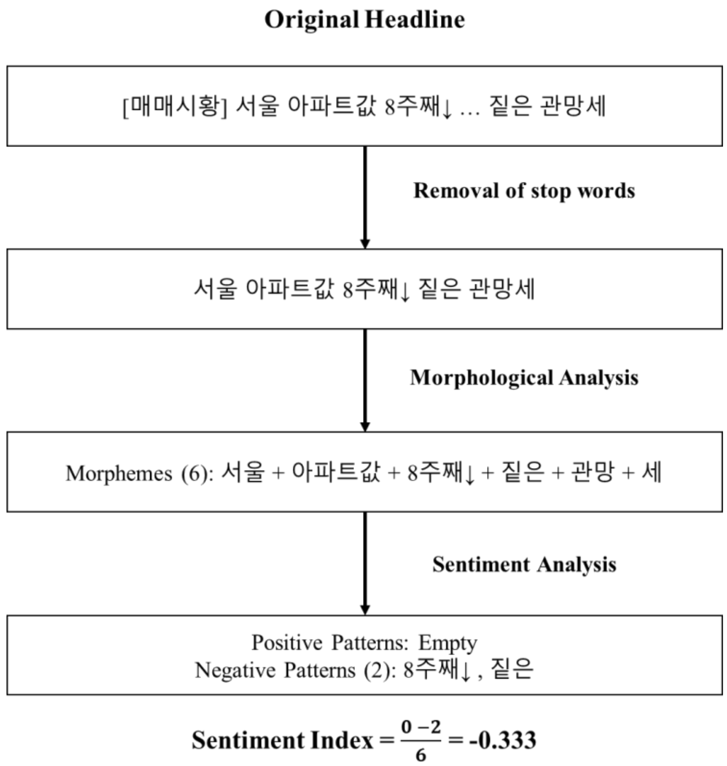

Figure 8. The Example of Sentiment Analysis

Time-series Analysis

To develop a predictive model, it is necessary to examine time-series trends of variables by drawing graphs. By doing so, it is possible to understand patterns of variables and guess the relationships among time-series variables.

To check the trends more quantitatively, the unit root test (Augmented Dickey-Fuller) is conducted to examine whether variables are stationary time-series or not. If a variable is non-stationary, the first differential values of variables should be considered in the time-series analysis.

And then, analysis of autocorrelation, cross-correlation and the Granger causality are conducted to build a predictive model. If a dependent variable has autocorrelations, a time-lagged dependent variable must be included in explanatory variables to solve the problem of white noise in the regression model. cross-correlations show a time lag which has the highest correlation between two variables. So, it gives information of time lags which are adequate for building the predictive model. Lastly, the Granger causality test is a statistical method for examining that one variable has a significant causal relationship with another variable.

Predictive Model

To predict the trends of the housing market, this study uses the sentiment index as the explanatory variable. In addition, the lagged values of the dependent variable should be included in the regression model because the time-series trends generally autocorrelated (Min and Choi, 2014). Therefore, the predictive models in this study are constructed as the Autoregressive and Distributed Lags (ARDL) models. A model of ARDL (1, q) is the equation:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 x_t + \cdots + \beta_q x_{t-q} + e_t \quad (2)$$

If we solve the equation (1), the equation becomes the model of infinite distributed lags. It means that the equation shows how much All lagged values of the explanatory variables (x) affect the dependent variables ($\frac{\beta_1 + \cdots + \beta_q}{1 - \alpha_1}$). However, there is a subjective process when q is determined. Akaike Information Criterion (AIC) and SIC (Schwarz Information Criterions) can be used to decide the value of q even though there are no thumb of rules. In other words, we can choose the value of q which makes the minimum values of AIC and SIC.

4 Results

4.1 Descriptive Statistics and Time-Series Patterns of Variables

Trends in News Articles

All variables are built as the monthly variables and summary statistics of all variables are shown in table 4. The temporal boundary is from July in 2011 to March in 2018. So, a total number of months is 81. However, HVT, NUA, and ARSPI can be collected until February in 2018.

The monthly average number of news articles related to the housing prices is approximately 15 and the average number per the press is 4.4. Actually, NAVER started to provide articles of Chosun Ilbo and JoongAng Ilbo since October in 2017 and April in 2017 respectively. So, the average number of news articles increased from that periods. The minimum number of published articles was 3 in May 2012, October 2013 and December 2013 while the maximum number was 97 in January 2018. At that times, many articles said that housing prices are increasing in especially "Gangnam" and criticize the housing policies established in 2017.

The average number of positive patterns in news articles is approximately 8.3 which is slightly higher than that of negative patterns. However, the average value of SI is the negative value (-0.005) and the distribution is positive skew. It means that there are more number of months which has the negative tone of news articles about the housing prices.

Trends of the Seoul Housing markets and Sentiment Index

In Seoul, 13,161 housings are traded every month and the lowest number of transactions was January 2013 (Figure 9). At that time, housing prices were also decreasing until the middle of 2013 (Figure 12 and 13). Since then, both the trends of housing volume of transaction and housing prices have been upwards although the trend of housing volume of transaction has fluctuated a lot. Thus, the highest number of transaction was August 2017 just before governments announced 8.2 Real Estate Policies for controlling overheating in the housing market. Despite the intention of the government, housing prices are still increasing until the end of study period.

Interestingly, the trends of Sentiment Index (SI) have preceded the trend of housing volume of transaction since January 2013. On the other hands, it is difficult to identify the relationship However, the trends of price indices are not independent to times, which means that the trends are not stationary. So, in this study, the 1st differential values of price indices are examined (Figure 20 and 21) and used in predictive models. The change rates of housing prices quite fluctuate, and the prices are likely to increase from every December to January. Moreover, the trends of SI and the change rates are similar, but it is difficult to identify whether SI precedes or not. So, it is required to conduct the more quantitative analysis.

In terms of unsold housings, the number of unsold housings is decreasing dramatically since the end of 2013 (Figure 10). So, there are only a few unsold housings in Seoul even though the average number is 1558. Furthermore, the trend is not stationary and the 1st differential value of the unsold housings should be examined (Figure 19). The graph of Figure 19 shows that the number of unsold housings in Seoul increased a lot in August 2012 and August 2013 when the housing volume of transaction decreased. However, there were little changes after the middle of 2016.

Comparing the Sentiment Index developed in this study with the consumer sentiment index developed in Korea Research Institute for Human Settlements (KRIHS), they have similar time-series patterns (Figure 15, 16 and 17). According to the report of KRIHS (2013), the consumer sentiment index can be used to predict the trends of trading volumes or housing prices. It implies that the Sentiment Index of the newspaper also can have the causal relationship with the trends of housing markets. Moreover, the Sentiment Index can be calculated more easily and less costly than the indices of KRIHS because KRIHS conduct the sample survey every month. Hence, news articles can represent the consumers' sentiment and it may be more efficient to develop sentiment index of the housing market by identifying tones of newspapers

Table 4. Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
The Number of Articles	81	15.358	12.955	3	97
Positive Pattern	81	8.333	9.278	0	55
Negative Pattern	81	7.025	6.399	0	43
The Number of Morpheme in the headlines	81	130.605	112.689	29	854
Sentiment Index	81	-0.005	0.070	-0.197	0.125
HVT	80	13161.090	5650.075	2451	24259
NUA	80	1558.300	1275.158	39	4331
BCH	81	7370.457	6490.707	1900	57046
ARSPI	80	140.666	15.693	122.8	182.9
ASPI	81	91.558	4.938	85.1	104.4
ACPI	81	86.600	9.930	71.2	100.7
CSIHM	81	121.459	11.789	93.7	143.7
CSIHSM	81	123.348	16.400	91.9	156.2
CSIHCM	81	119.572	12.460	95.5	144.8
D_HVT	79	123.595	3591.407	-12584	8148
D_UA	79	-22.506	293.354	-565	1448
D_BCH	80	14.663	8852.336	-48689	48137
D_ARSPI	79	0.589	1.246	-1.7	4.5
D_ASPI	80	0.134	0.443	-0.8	1.4
D_ACPI	80	0.366	0.392	-0.4	1.6
D_CSIHM	80	-0.113	7.958	-19.4	21.1
D_CSIHSM	80	0.276	8.674	-32.9	17.2
D_CSIHCM	80	-0.501	9.113	-21.4	24.9

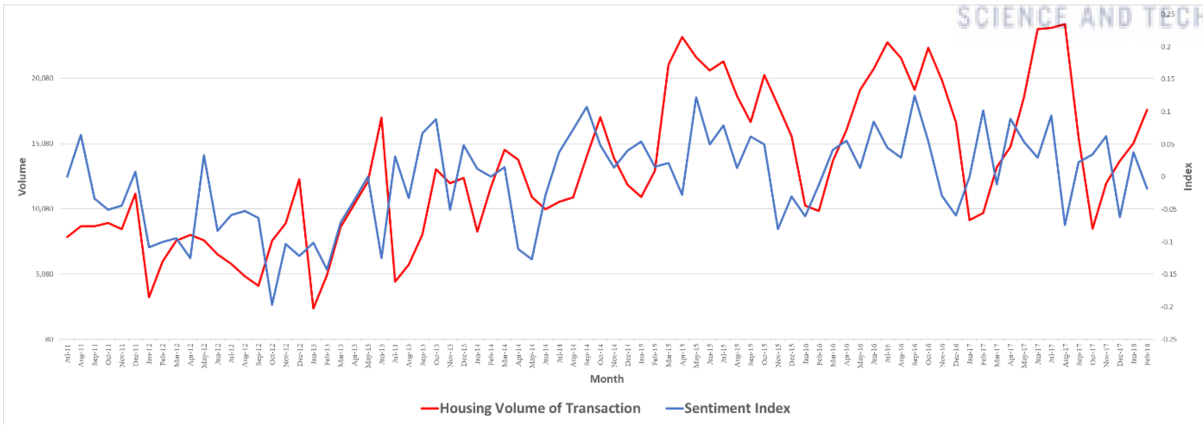

Figure 9. Trends of Housing Volume of Transaction and Sentiment Index, 2011.7 - 2018.2

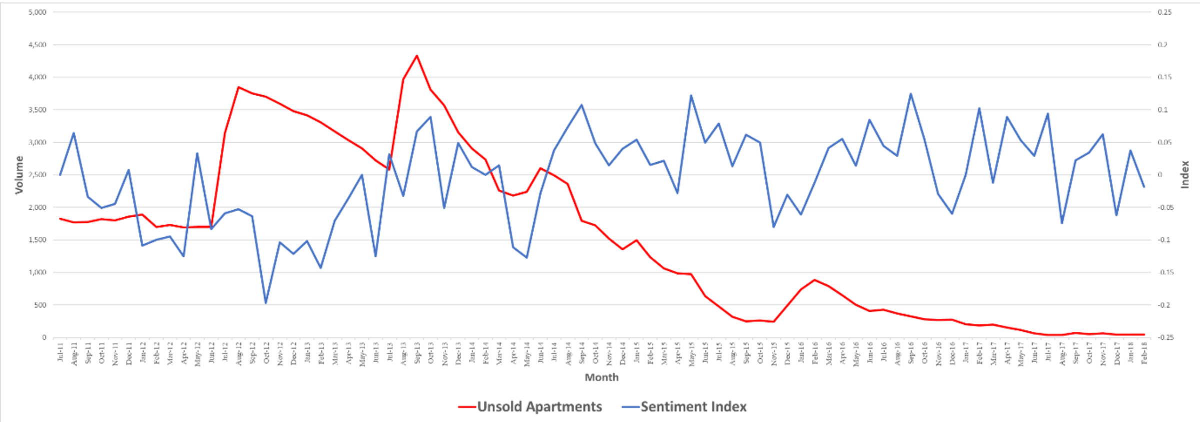

Figure 10. Trends of Unsold Apartments and Sentiment Index, 2011.7 - 2018.2

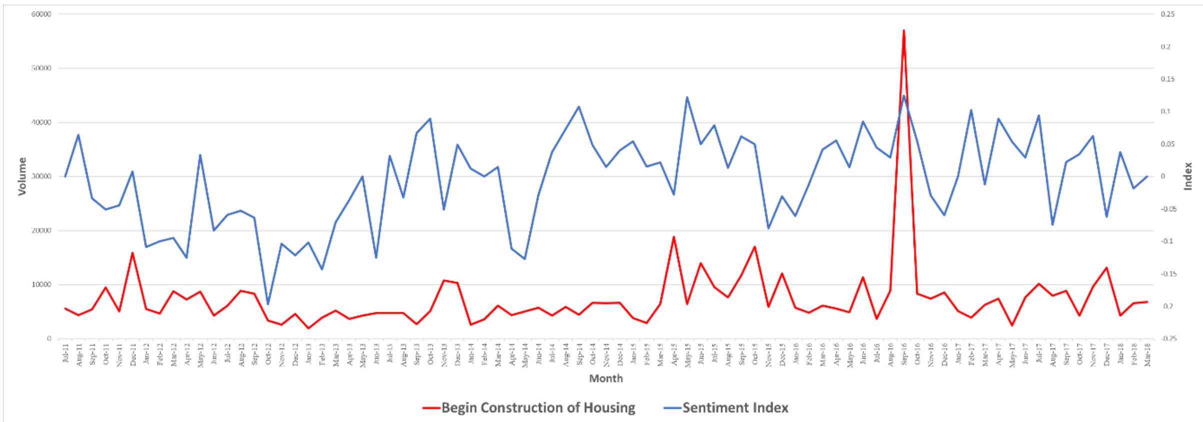

Figure 11. Trends of Begin Construction of Housings and Sentiment Index, 2011.7 - 2018.3

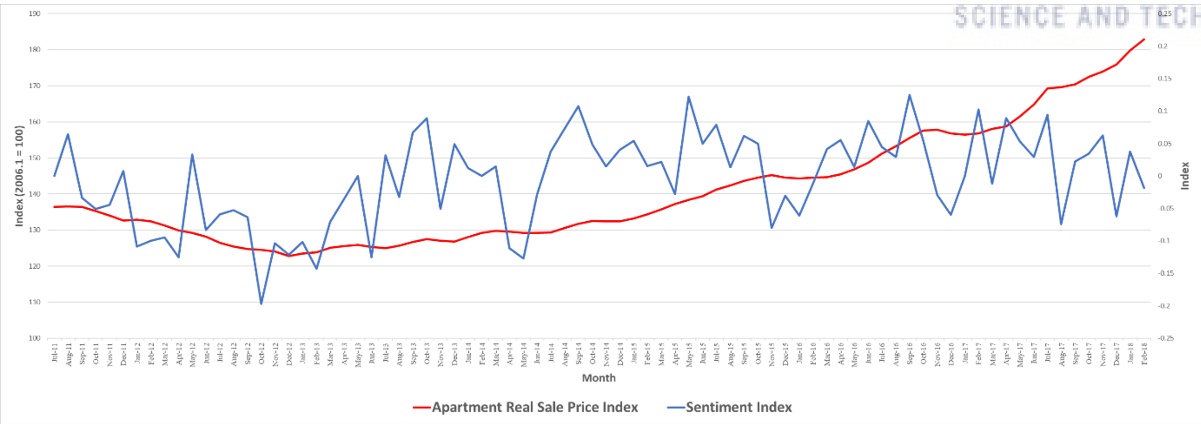

Figure 12. Trends of Apartment Real Sale Price Index and Sentiment Index, 2011.7 - 2018.2

Figure 13. Trends of Apartment Sale Price Index and Sentiment Index, 2011.7 - 2018.3

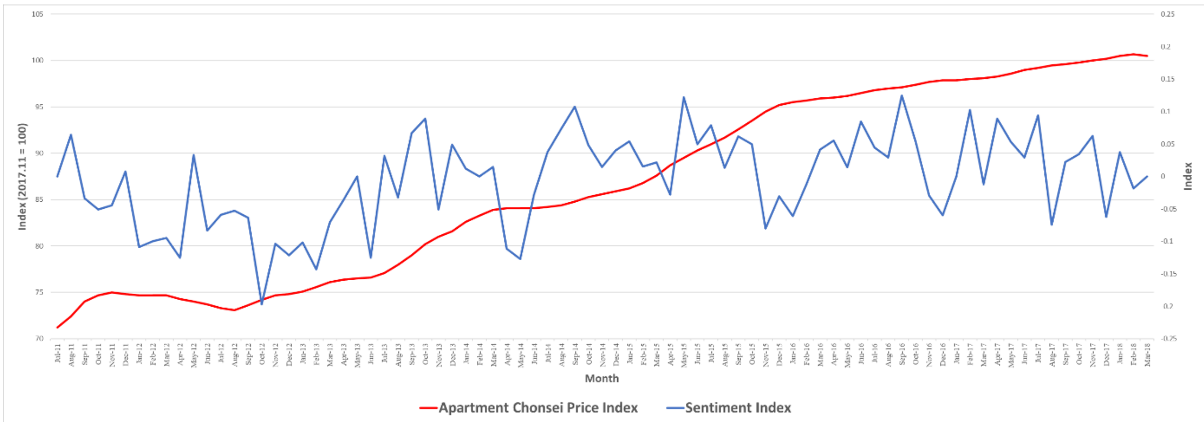

Figure 14. Trends of Apartment Chonsei Price Index and Sentiment Index, 2011.7 - 2018.3

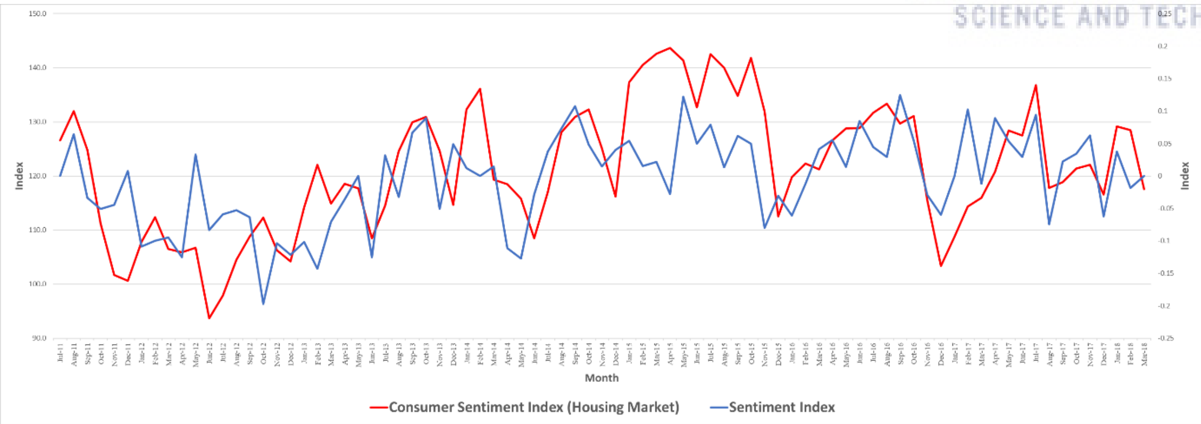

Figure 15. Trends of Consumer Sentiment Index of Housing Market and Sentiment Index, 2011.7 - 2018.3

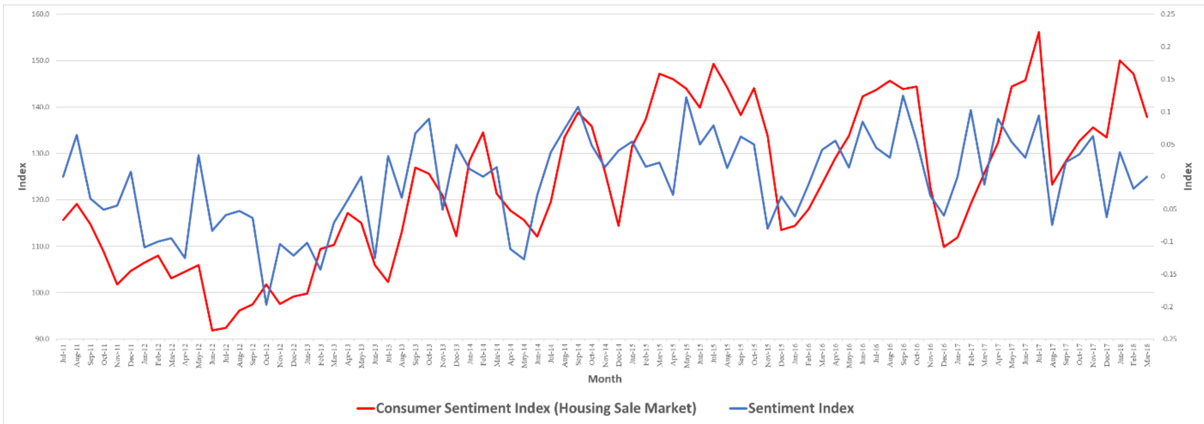

Figure 16. Trends of Consumer Sentiment Index of Housing Sale Market and Sentiment Index, 2011.7 - 2018.3

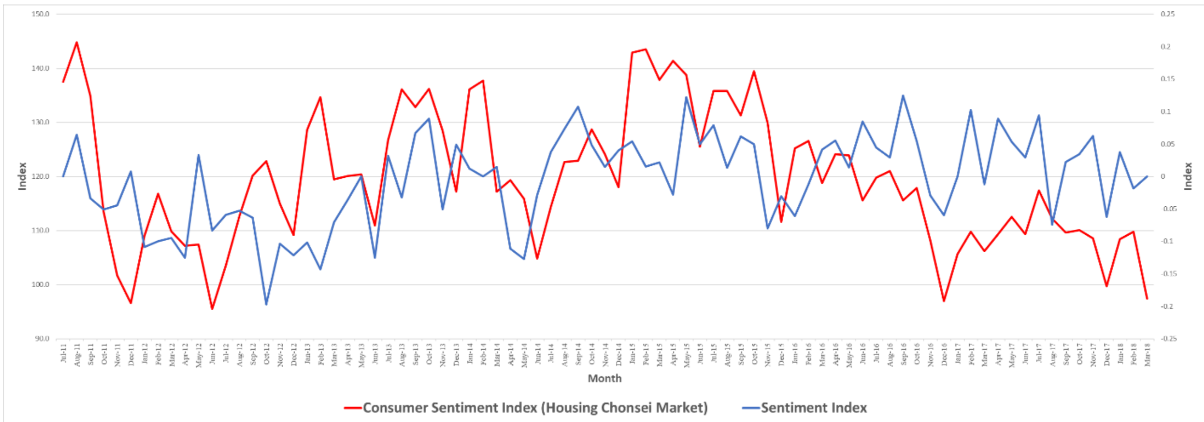

Figure 17. Trends of Consumer Sentiment Index of Housing Chonsei Market and Sentiment Index, 2011.7 - 2018.3

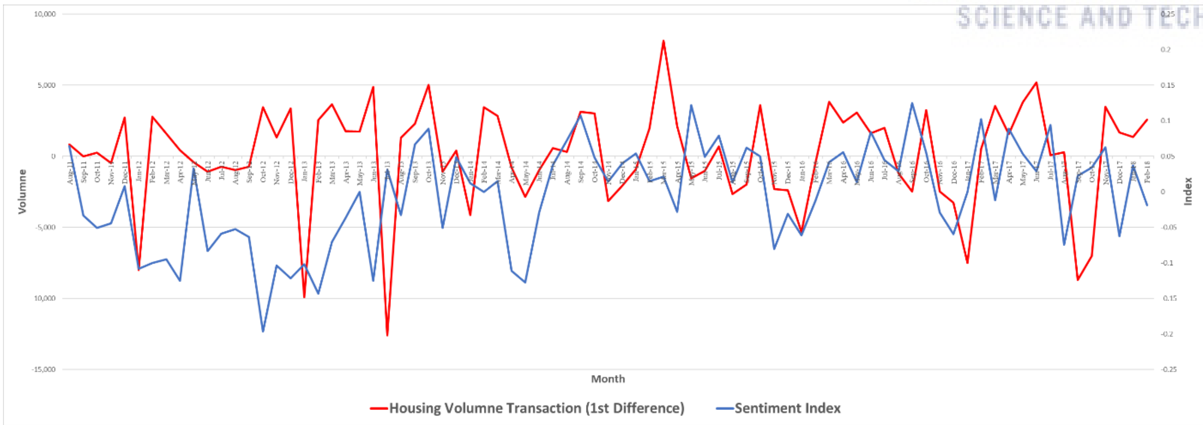

Figure 18. Trends of Housing Volume Transaction (1st Difference) and Sentiment Index, 2011.7 - 2018.2

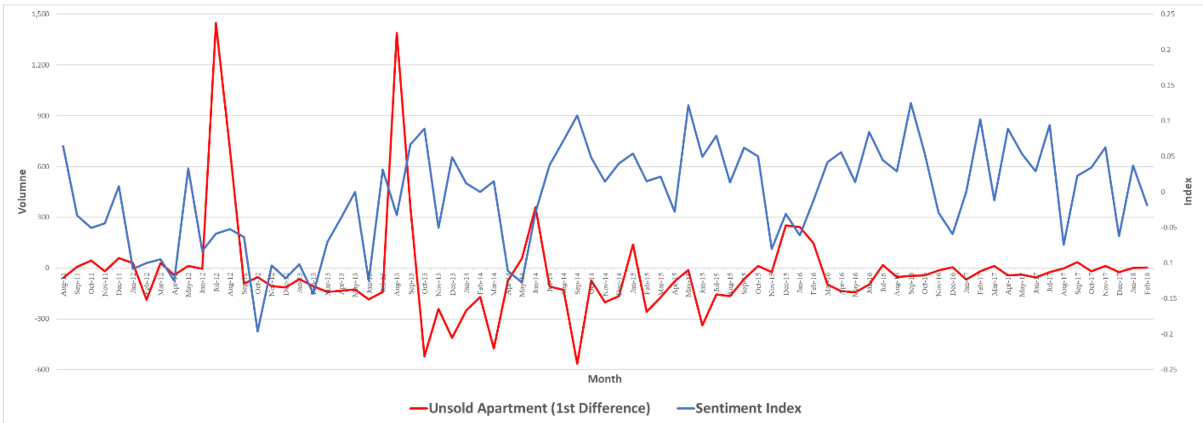

Figure 19. Trends of Unsold Apartment (1st Difference) and Sentiment Index, 2011.7 - 2018.2

Figure 20. Trends of Apartment Real Sale Price Index (1st Difference) and Sentiment Index, 2011.7 - 2018.2

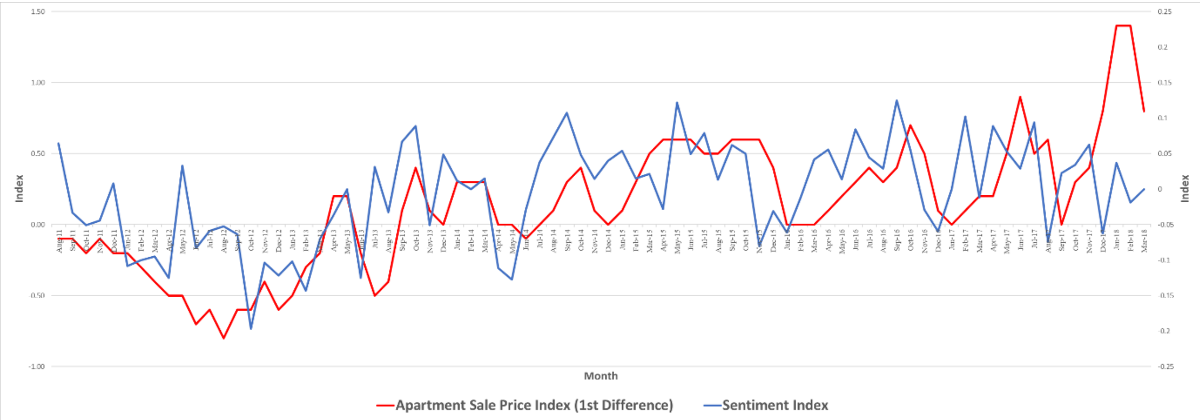

Figure 21. Trends of Apartment Sale Price Index (1st Difference) and Sentiment Index, 2011.7 - 2018.3

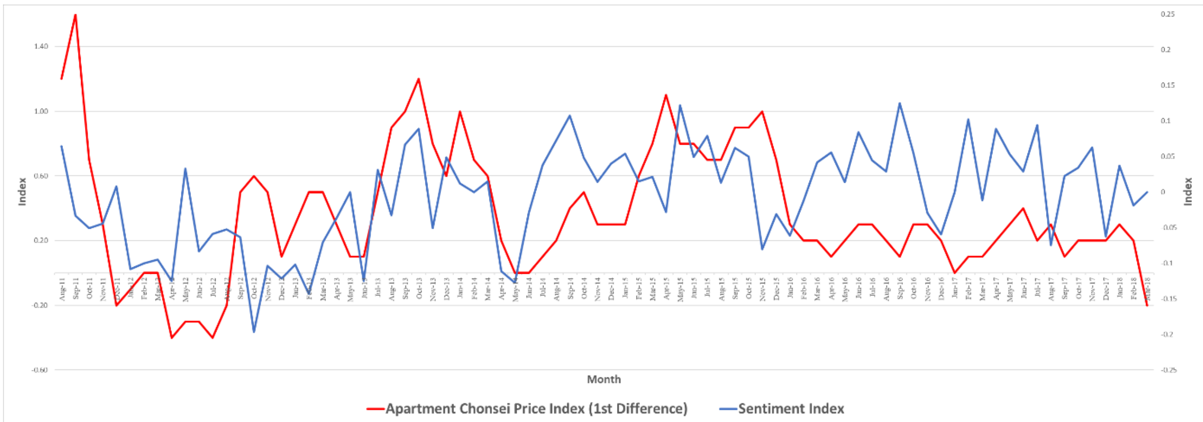

Figure 22. Trends of Apartment Chonsei Price Index (1st Difference) and Sentiment Index, 2011.7 - 2018.3

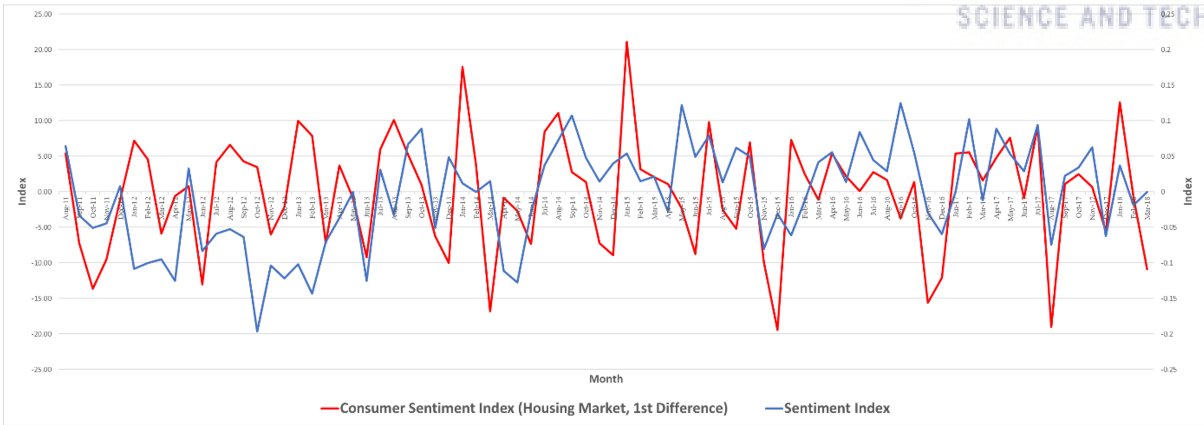

Figure 23. Trends of Consumer Sentiment Index of Housing Market (1st Difference) and Sentiment Index, 2011.7 - 2018.3

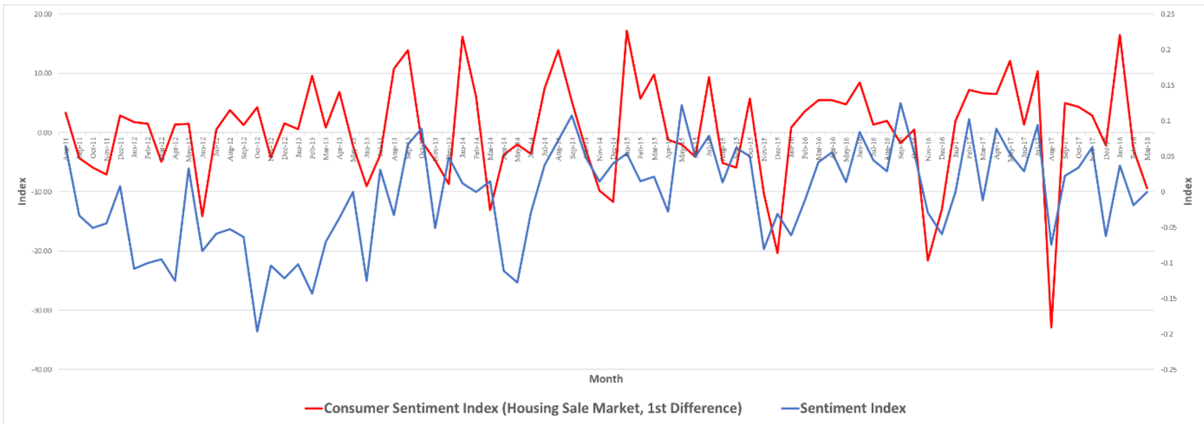

Figure 24. Trends of Consumer Sentiment Index of Housing Sale Market (1st Difference) and Sentiment Index, 2011.7 - 2018.3

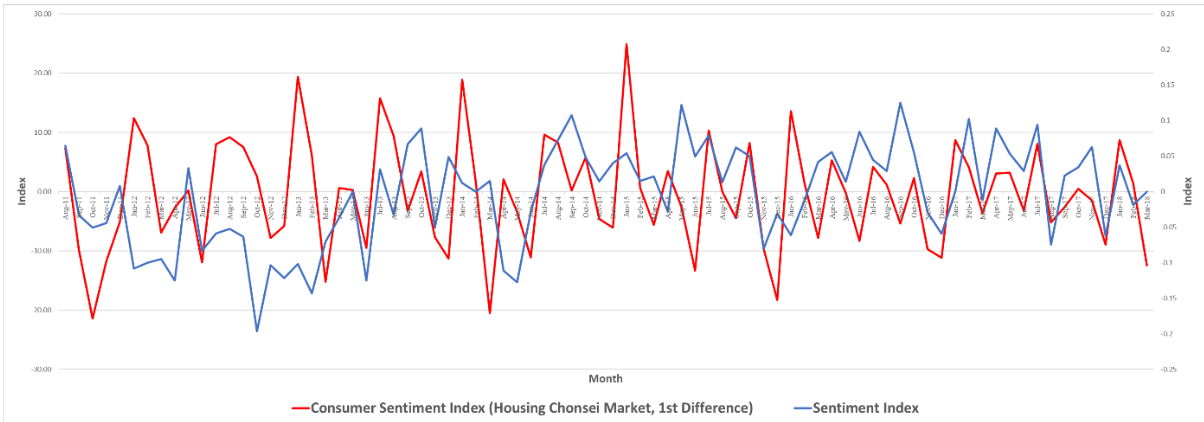

Figure 25. Trends of Consumer Sentiment Index of Housing Chonsei Market (1st Difference) and Sentiment Index, 2011.7 - 2018.3

4.2 The Results of Time-Series Analysis

Autocorrelation of Variables

The results of autocorrelations and partial autocorrelations are shown in Table 5. Except for BCH, all variables are autocorrelated. So, most variables in the housing market have some repeating patterns over times. Therefore, time-lagged dependent variables should be included in explanatory variables for developing predictive models when variables are autocorrelated. It may be taken for granted that housing prices or the number of unsold housings in last month affect ones in this month. However, the amount of construction is not autocorrelated.

Table 5. The results of autocorrelations and partial autocorrelations

Variables	Autocorrelations	Partial Autocorrelations
SI	O (Maximum 3 month)	O (Maximum 2 month)
HVT	O (Maximum 3 month)	O (Maximum 1 month)
NUA	O (Maximum 7 month)	O (Maximum 2 month)
BCH	X	X
ARSPI	O (Maximum 7 month)	O (Maximum 2 month)
ASPI	O (Maximum 7 month)	O (Maximum 3 month)
ACPI	O (Maximum 8 month)	O (Maximum 2 month)
CSIHM	O (Maximum 3 month)	O (Maximum 1 month)
CSIHSM	O (Maximum 4 month)	O (Maximum 1 month)
CSIHCM	O (Maximum 3 month)	O (Maximum 1 month)

Cross-Correlation of Variables

The results of cross-correlation analysis between variables of housing markets and Sentiment Index are shown in From Figure 26 to Figure 34. Most variables of housing markets are correlated with time-lagged SI and the relationships usually match with expected ones.

The Housing Volume of Transaction (HVT)

The housing volume of transaction has the highest coefficient of correlation (0.589) with one-month lagged SI (Figure 25). It means that the tone of newspapers about the housing prices in last month has the strongest effect on the trading volume of housing in Seoul. Moreover, the effect of the sentiment of newspapers continues for three months according to Figure 5. Meanwhile, the change rate of housing volume of transaction is less correlated with SI, but the highest coefficient of correlation (0.232) with also one-month lagged SI (Figure 26). In summary, both the housing volume of transaction and the change rate of one seem to be affected by the sentiment of newspapers.

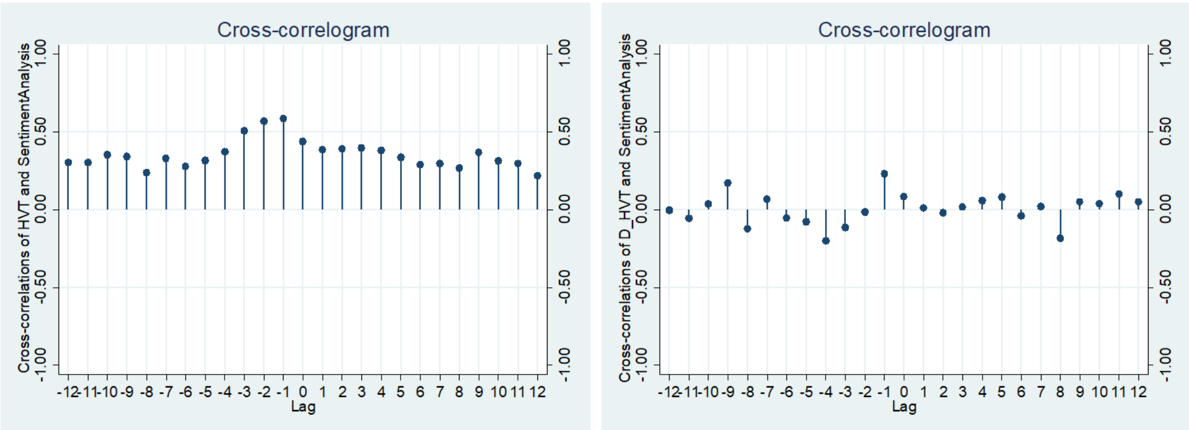

Figure 26. The Cross-Correlation between of Housing Volume of Transaction and Sentiment Index: Left - HVT and Right - D_HVT

The Number of Unsold Housings (NUH)

The results of cross-correlation analysis between the number of unsold housings in Seoul and SI are shown in Figure 27. Compared to HVT, the effect of SI on NUH lasts for longer periods. For example, the coefficient of correlation between SI a year ago and NUH this month is still higher than 0.5. On the other hands, there may be no causal relationship between SI and the change rates of NUM.

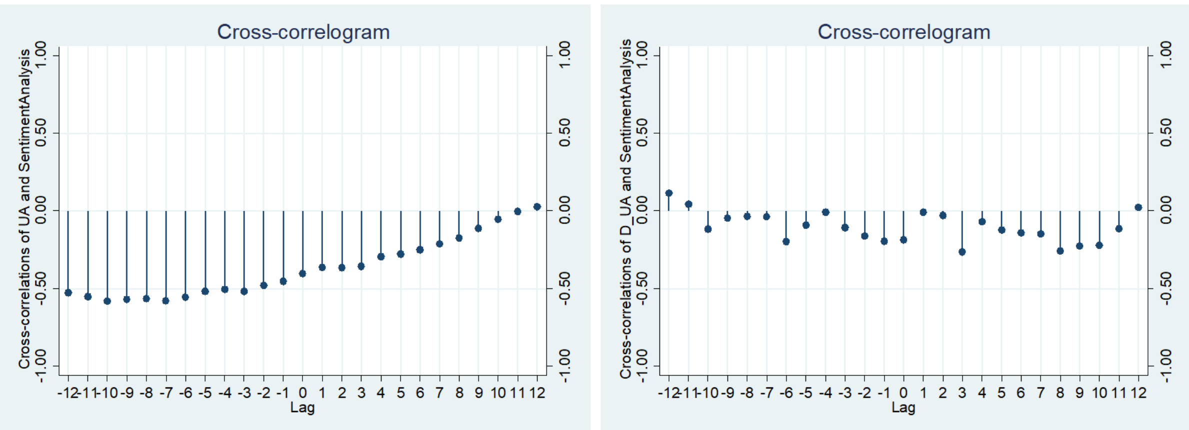

Figure 27. The Cross-Correlation between of the Number of Unsold Housings and Sentiment Index: Left - NUT and Right - D_HVT

The Number of Begin Construction of Housings (BCH)

The results of cross-correlation analysis between the number of begin construction of Housings in Seoul and SI are shown in Figure 28. Compared to other trends, the relationship between SI on NUH is relatively weak. It may be because the construction companies consider future's market rather than present's because begin construction is the just start point and they can adjust a time of completion. At least, whether housing prices are anticipated positively or negatively is not a significant to start the construction.

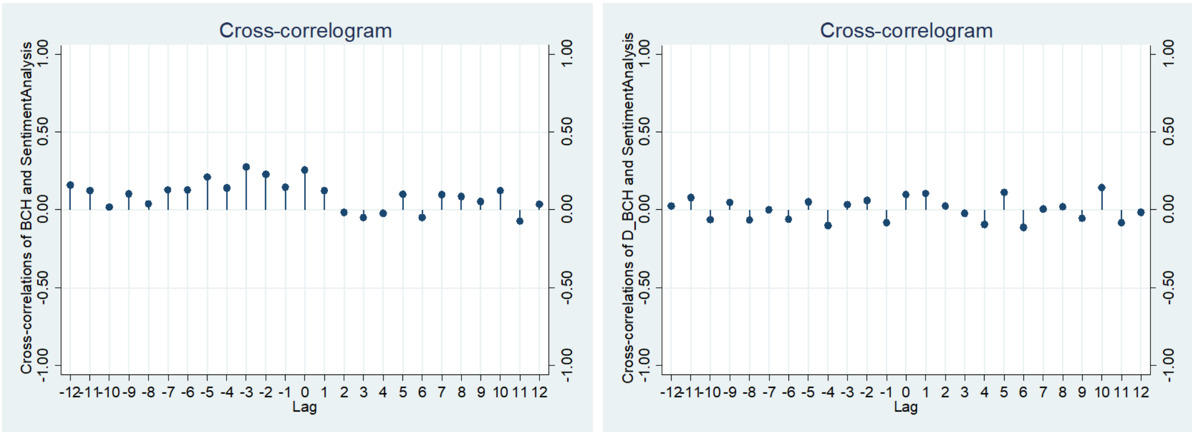

Figure 28. The Cross-Correlation between of Begin Construction of Housings and Sentiment Index: Left - BCH and Right - D_BCH

The Housing Price Indices: Apartment Real Sale Price Index (ARSPI), Apartment Sale Price Index (ASPI) and Apartment Chonsei Price Index (ACPI)

The results of cross-correlation analysis between the housing price indices in Seoul and SI are shown in Figure 29,30 and 31. The pattern of cross-correlation is similar to the case of NUH, but the coefficient is positive. In other words, if the newspapers of housing prices are written negatively, the housing prices tend to decrease one or two month later.

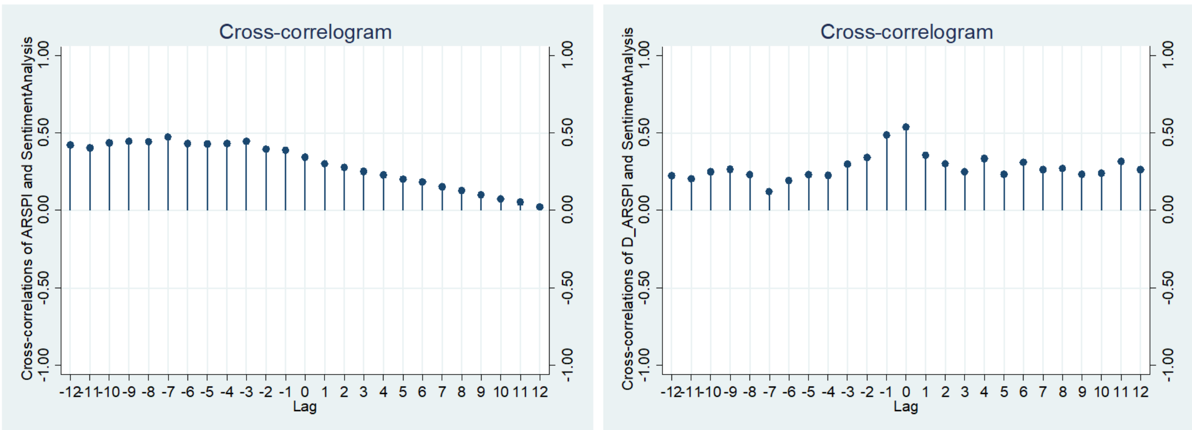

Figure 29. The Cross-Correlation between of Apartment Real Sale Price Index and Sentiment Index: Left - ARSPI and Right - D_ARSPI

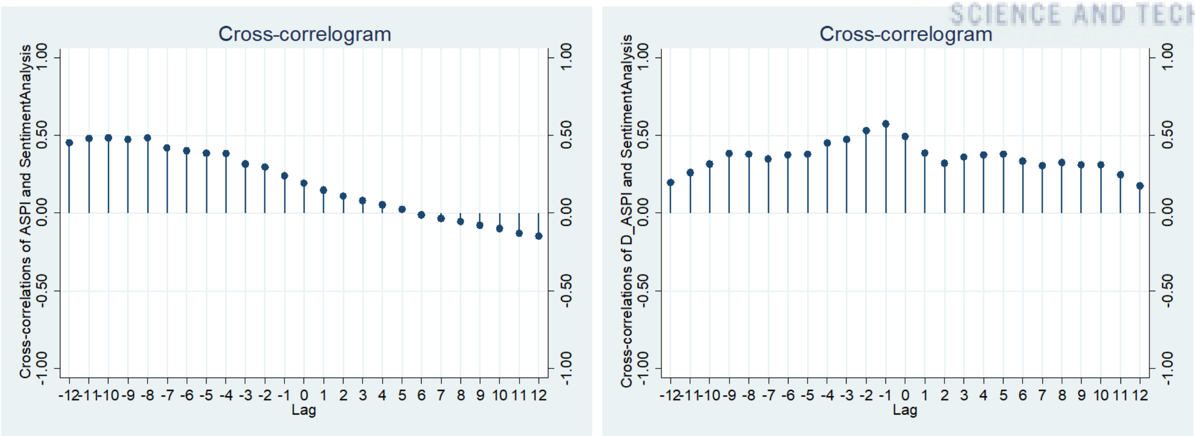

Figure 30. The Cross-Correlation between of Apartment Sale Price Index and Sentiment Index: Left - ASPI and Right - D_ASPI

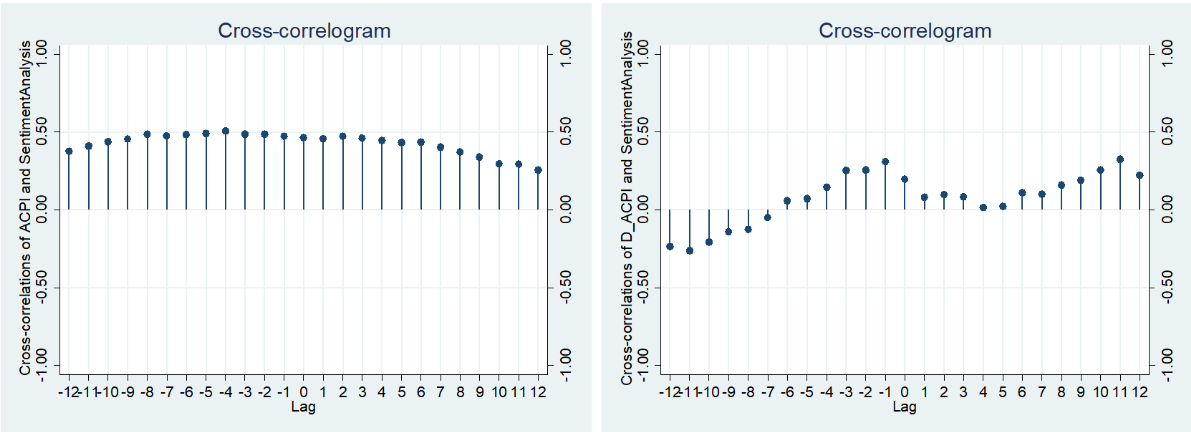

Figure 31. The Cross-Correlation between of Apartment Chonsei Price Index and Sentiment Index: Left - ACPI and Right - D_ACPI

The Consumer Sentiment Index of Housing market (CSIHM), Housing Sale Market (CSIHSM) and Housing Chonsei Market (CSIHCM)

The results of cross-correlation analysis between the consumer sentiment indices in Seoul and SI are shown in Figure 32, 33 and 34. It seems that the consumer sentiment indices and SI have similar patterns. It means that when SI or the consumer sentiment indices increases, others tend to increase at the same times. Thus, SI has some potential for substitution of the consumer sentiment indices.

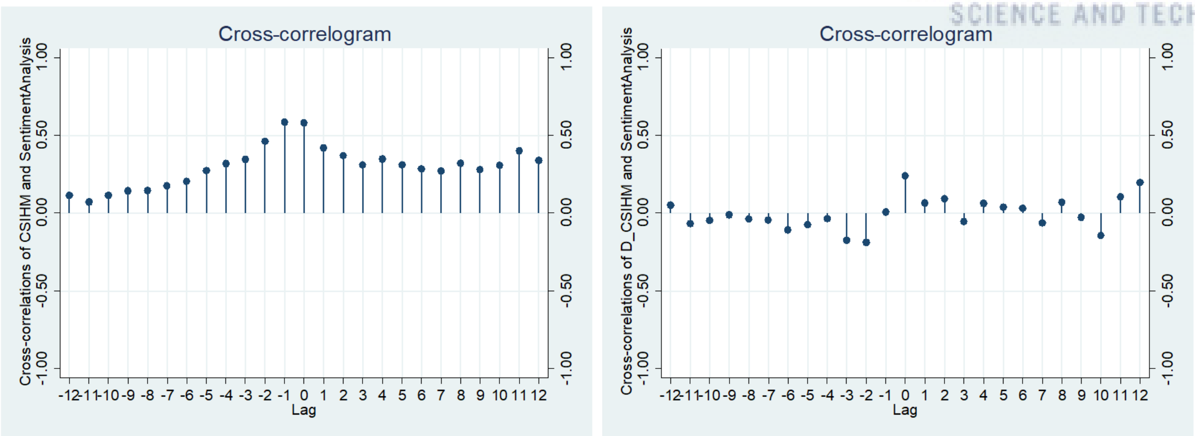

Figure 32. The Cross-Correlation between of Consumer Sentiment Index of Housing Market and Sentiment Index: Left - CSIHM and Right - D_CSIHM

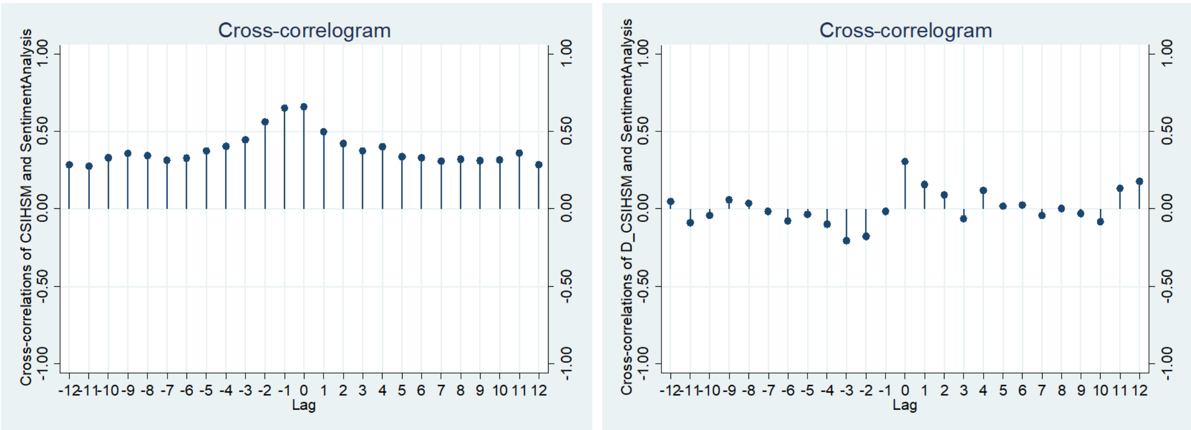

Figure 33. The Cross-Correlation between of Consumer Sentiment Index of Housing Sale Market and Sentiment Index: Left - CSIHSM and Right - D_CSIHSM

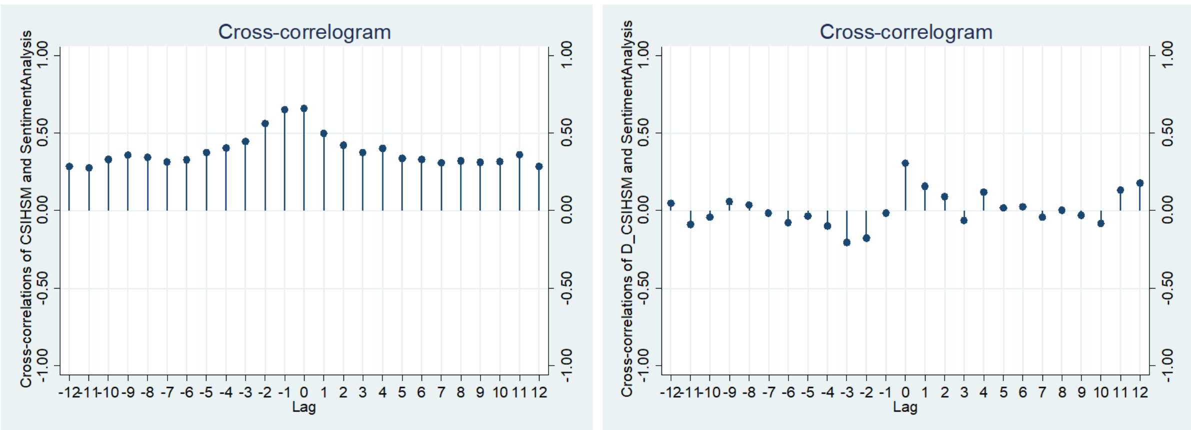

Figure 34. The Cross-Correlation between of Consumer Sentiment Index of Housing Chonsei Market and Sentiment Index: Left - CSIHCM and Right - D_CSIHCM

The results of Granger causality

The results of Granger causality test are shown in table 6. In summary, the Sentiment Index (SI) has the causal relationships with Apartment Real Sale Price Index (ARSPI), Apartment Sale Price Index (ASPI) and the change rates of Housing Volume of Transaction (HVT) and Apartment Chonsei Price Index (ACPI). In other words, the past values of SI can used to predict those variables. Meanwhile, it is shown that both SI and two consumer sentiment Indexes are correlated, but there are no causal relationships. So, the newly developed SI has a high possibility of similar function of existing indices.

Table 6. The results of granger causality analysis

Variables 1	Variables 2	Granger Causality	Granger Causality (1st difference)
HVT	Sentiment Index (SI)	↔	←
NUA		↔	→
BCH		←	X
ARSPI		←	X
ASPI		←	↔
ACPI		↔	←
CSIHM		↔	X
CSIHSM		↔	X
CSIHCM		X	X

4.3 The Results of Predictive Modellings: Autoregressive and Distributed Lags (ARDL)

Based on the time-series analysis, the final predictive modellings are constructed. The results are shown in Table 7. If a dependent variable is autocorrelated, the one-month lagged value of the dependent variable is included in the explanatory variables. Moreover, 1st differential values are included in the dependent variables because some variables are not stationary is not stationary,

Some variables such as BCH, D_HVT, D_NUA, D_ACPI, D_CSIHM, D_CSIHSM and D_CSIHCM are failed to predict by using the Sentiment Index (SI), other variables such as HVT, NUA, CSIHM, CSIHSM, D_ASPI, D_ACPI have high explanatory powers in the predictive modellings. For instance, the all past values of SI have effect on the trading volumes in this month positively while they have negative effect on the number of unsold housings. It has good performance to predict the change rates of Apartment Sale Price Index (ASPI).

Table 7. The results of predictive models

DEPENDENT VARIABLE	HVT	HVT	NUA	NUA	BCH
Lagged D.V	-	0.621	-	0.944	-
Long-term effect of SI	64378.530	76810.546	-11709.626	-28787.394	32568.357
Lagged SI (1 month)	33206.950	21919.000	-5377.366	-898.833	170.227
Lagged SI (2 month)	31171.580	7202.000	-6332.260	-726.740	12186.110
Lagged SI (3 month)	-	-	-	-	20212.020
Constant	13614.760	5254.000	1491.841	58.550	7625.299
R-squared	0.460	0.711	0.284	0.951	0.054
N	78	78	78	78	78
DEPENDENT VARIABLE	**CSIHM**	**CSIHM**	**CSIHSM**	**CSIHSM**	**-**
Lagged D.V	-	0.654	-	0.768	-
Long-term effect of SI	98.738	100.323	152.255	145.753	-
Lagged SI (1 month)	98.738	34.708	152.255	33.825	-
Lagged SI (2 month)	-	-	-	-	-
Lagged SI (3 month)	-	-	-	-	-
Constant	121.889	42.099	124.205	29.029	-
R-squared	0.337	0.619	0.417	0.747	-
N	80	80	80	80	-

(Continuned)

DEPENDENT VARIABLE	CSIHCM	CSIHCM	D_HVT	D_NUA	D_NUA
Lagged D.V	-	0.711	-	-	0.266
Long-term effect of SI	45.223	53.782	3065.541	-1022.066	-732.659
Lagged SI (1 month)	45.223	15.559	18549.700	-630.803	-497.827
Lagged SI (2 month)	-	-	-3563.969	-391.262	-234.833
Lagged SI (3 month)	-	-	-11920.190	-	-
Constant	119.574	34.249	147.206	-27.229	-19.667
R-squared	0.054	0.524	0.0637	0.0195	0.0761
N	80	80	77	78	78

DEPENDENT VARIABLE	D_ARSPI	D_ARSPI	D_ASPI	D_ASPI	-
Lagged D.V	-	0.672	-	0.787	-
Long-term effect of SI	8.549	6.481	3.607	1.174	-
Lagged SI (1 month)	8.549	2.125	3.607	1.174	-
Lagged SI (2 month)	-	-	-	-	-
Lagged SI (3 month)	-	-	-	-	-
Constant	0.630	0.230	0.152	0.042	-
R-squared	0.2268	0.5199	0.3206	0.7756	-
N	79	79	80	80	-

(Continued)

DEPENDENT VARIABLE	D_ACPI	D_ACPI	D_CSIHM	D_CSIHSM	D_CSIHCM
Lagged D.V	-	0.768	-	-	-
Long-term effect of SI	2.186	2.933	-20.378	-25.391	-15.433
Lagged SI (1 month)	1.356	1.017	19.386	17.298	21.173
Lagged SI (2 month)	0.831	-0.336	-21.393	-18.582	-23.893
Lagged SI (3 month)	-	-	-18.371	-24.107	-12.713
Constant	0.367	0.072	-0.180	0.177	-0.540
R-squared	0.0981	0.6546	0.0299	0.0274	0.0087
N	79	79	78	78	78

5 Conclusion

This research introduces text-mining approach to housing market analysis. It develops Sentiment Index which quantifies the tones of newspapers regarding housing prices. And then, this study explores the trends of Sentiment Index and various variables of housing markets in Seoul. The results show that the tones of newspapers are potential sources to predict several trends of the housing markets. In other words, the results support research hypothesis that news articles can affect the decision-making process and it influences the trends of the housing market.

However, it is hard to say that there are strong causal relationships because Sentiment Index and the trends of housing markets are usually affected by each other. The one of the main functions of newspaper is to deliver the objective information such as the range of change in housing prices. So, it is also possible that the news articles reflect the past trends. In the results, Sentiment Index were the preceding index in some periods, otherwise it is not preceding.

This study contributes to housing studies by introducing some techniques of text-mining and identifying. It is a meaningful because text-mining of Korean is at a developing level and there are few studies applying it for the housing study. Therefore, it is expected that it becomes possible to extract more significant information from text data as the technique develops in the future.

REFERENCES

1. Adams, F. G. (1964). Consumer Attitudes, Buying Plans, and Purchases of Durable Goods: A Principal Components, Time Series Approach, The MIT Press, pp. 347-355.

2. Balahur, A., & Steinberger, R. (2009). Rethinking Sentiment Analysis in the News: from Theory to Practice and back.

3. Carroll, C. D., Fuhrer, J. C. and Wilcox, D. W. (1994). Does Consumer Sentiment Forecast Household Spending? If So, Why?, The American Economic Review, 84(5), pp. 1397-1408.

4. Choi, H. and Rhim, B. (2009). Role of the Housing Price Forecast in Housing Price and Business Cycle, The Korea Spatial Planning Review, pp. 141-158.

5. Engelhardt, G. V. (2003). Nominal Loss Aversion, Housing Equity Constraints, and Household Mobility: Evidence from the United States, Journal of Urban Economics, 53(1), pp.171-195.

6. Fama, E. F. (1965). Random walks in stock market prices. Financial Analysts Journal, 21, pp.55–59.

7. Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25, pp.383–417.

8. Fairclough, N. (2000). Discourse, Social Theory and Social Research: The Discourse of Welfare Reform, Journal of Sociolinguistics, 4(2), pp. 163–195.
9. Friesen, G., and Weller, P. A. (2006). Quantifying cognitive biases in analyst earnings forecasts. Journal of Financial Markets, 9, pp. 333–365.
10. Hearst, M. (1999). Untangling Text Data Mining, in the Proceedings of ACL'99: the annual Meeting of the Association for Computational Linguistics, University of Maryland, 20-26 June.
11. Kim, D. W. and Yu, J. S. (2013). An Analysis on How Psychological Attitudes on the House Price Affect the Trading Volume, Housing Studies Review, 21(2), pp.73-92.
12. Kam, M. and Song, M. (2012). A Study on Differences of Contents and Tones of Arguments among Newspapers Using Text Mining Analysis, Journal of Intelligence and Information Systems, 18(3), pp. 53~77.
13. Kim, J. H. and Choi, Y. Y. (2016). A Study on the Impact of Property Consumer Sentiment on Apartment Transaction, Korea Real Estate Review, 26(1). pp.7-17.
14. Kim S. W. and Kim N. G. (2014). A Study on the Effect of Using Sentiment Lexicon in Opinion Classification, Journal of Intelligent Information Systems, 20(1), pp.133-148.
15. Kim, Y. S., Kim, N. and Jeong, S. R. (2012). Stock-Index Invest Model Using News Big Data Opinion Mining, Journal of Intelligence and Information Systems, 18(2), pp.143-156

16. Korea Research Institute for Human Settlements. (2013). True & Fair Housing Market Report, 2013.04.22.
17. Lee, S. H., Cui, J. and Kim, J. W. (2016). Sentiment Analysis on Movie Review through Building Modified Sentiment Dictionary by Movie Genre," Journal of Intelligence and Information Systems, 22(2), pp. 97~113.
18. Min, I. S. and Choi, P. S. (2014). STATA Time Series Data Analysis, The Korean Journal of STATA.
19. Munro, M. (2018). House Price Inflation in the News: a Critical Discourse Analysis of Newspaper Coverage in the UK, Housnig Studies, DOI: 10.1080/02673037.2017.1421911.
20. Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. Expert Systems with Applications, 41(16), pp.7653-7670.
21. Nyman, R., Kapadia, S., Tuckett, D., Gregory, D., Ormerod, P., & Smith, R. (2018). News and narratives in financial systems: exploiting big data for systemic risk assessment. Staff Working Paper No. 704. Bank of England.
22. Richardson, J. E. (2007). Analysing Newspapers: An Approach from Critical Discourse Analysis. (NewYork: Palgrave Macmillan).
23. Robertson, C., Geva, S., & Wolff, R. (2006). What types of events provide the strongest evidence that the stock market is affected by company specific news? Proceedings of the fifth Australasian conference on data mining and analytics, 61, pp.145-153. Sydney, Australia: Australian Computer Society Inc.
24. Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. Decision Support Systems.
25. Seo, B. J., Lee, H. J. and Jung, C. M. (2010). An Empirical Study on the Unsold New Apartment of Regional housing Market using Partial Adjustment and Dynamic Panel model, Journal of Korea Planning Association, 45(7), pp.97~117
26. Wisniewski, T. P., & Lambe, B. (2013). The role of media in the credit crunch: The case of the banking sector. Journal of Economic Behavior and Organization, 85, pp.163–175.
27. Yu, L. C., Wu, J. L., Chang, P. C., & Chu, H. S. (2013). Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. Knowledge-Based Systems.